Nonlinear regression model

Statistics and Big Data

Niccolò Salvini, PhD

UCSC

Academic Year 2025-2026

Course: Statistics and Big Data

Overview

What is Logistic Regression?

How do we predict outcomes that are categorical, such as whether a mouse is obese or not?

What is Logistic Regression?

How do we predict outcomes that are categorical, such as whether a mouse is obese or not? What if we could model the probability of an event occurring rather than predicting a continuous value?

Linear Regression Recap

Let's consider a familiar example: linear regression. Imagine we have data on mouse weight and size. By fitting a line to this data, we can:

1. Calculate the correlation (R-squared).
2. Determine statistical significance (p-value).
3. Predict size based on weight.

Linear Regression Recap

Let's consider a familiar example: linear regression. Imagine we have data on mouse weight and size. By fitting a line to this data, we can:

1. Calculate the correlation (R-squared).
2. Determine statistical significance (p-value).
3. Predict size based on weight.

This process illustrates how linear regression serves as a foundational machine learning technique.

Transition to Logistic Regression

Now that we understand linear regression, how does logistic regression differ?

Transition to Logistic Regression

Now that we understand linear regression, how does logistic regression differ? Instead of predicting a continuous outcome, logistic regression predicts a binary outcome: true or false, obese or not obese.

Core Concept of Logistic Regression

Logistic regression fits an S-shaped curve, known as the logistic function, to the data.

Core Concept of Logistic Regression

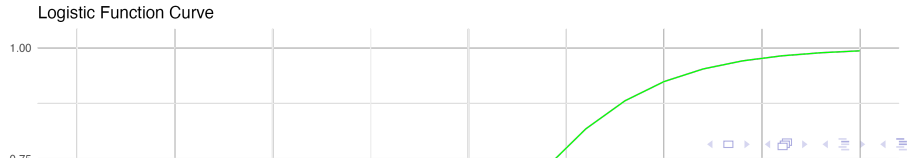
Logistic regression fits an S-shaped curve, known as the logistic function, to the data. This curve represents the probability of an event occurring, ranging from 0 to 1.

Mathematical Formulation

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}$$

Where:

- $P(Y = 1|X)$ is the probability of the event occurring.
- β_0 is the intercept.
- β_1 is the coefficient for the predictor variable X.



Interpreting the Logistic Function

What does the logistic function tell us about mouse obesity?

Interpreting the Logistic Function

What does the logistic function tell us about mouse obesity?

- A heavy mouse has a high probability of being obese.
- An intermediate weight mouse has a 50% chance.
- A light mouse has a low probability.

This probabilistic interpretation is crucial for classification.

Classification with Logistic Regression

How do we classify a mouse as obese or not?

Classification with Logistic Regression

How do we classify a mouse as obese or not? If the predicted probability exceeds 50%, we classify it as obese; otherwise, it is classified as not obese.

Classification with Logistic Regression

How do we classify a mouse as obese or not? If the predicted probability exceeds 50%, we classify it as obese; otherwise, it is classified as not obese. This classification process can be extended to multiple predictors, such as weight, genotype, and age.

Testing Variable Significance

How do we determine which variables are useful in our model?

Testing Variable Significance

How do we determine which variables are useful in our model? We utilize Wald's test to assess whether the effect of each variable is significantly different from zero.

Mathematical Formulation

$$W = \frac{\hat{\beta}}{\text{SE}(\hat{\beta})}$$

Where:

- $\hat{\beta}$ is the estimated coefficient.
- $\text{SE}(\hat{\beta})$ is the standard error of the coefficient.

This helps us identify non-contributing variables, such as astrological sign.

Maximum Likelihood Estimation

What is the method used to fit the logistic regression model?

Maximum Likelihood Estimation

What is the method used to fit the logistic regression model? Unlike linear regression, logistic regression employs maximum likelihood estimation (MLE).

Maximum Likelihood Estimation

What is the method used to fit the logistic regression model? Unlike linear regression, logistic regression employs maximum likelihood estimation (MLE).

Mathematical Formulation

$$L(\beta) = \prod_{i=1}^{n} P(Y_i|X_i)^{Y_i}(1 - P(Y_i|X_i))^{1-Y_i}$$

Where:

- $L(\beta)$ is the likelihood function.
- Y_i is the observed outcome.

This method finds the parameter estimates that maximize the likelihood of the observed data.

Summary of Logistic Regression

In summary, logistic regression allows us to classify outcomes using both continuous and categorical predictors.

Summary of Logistic Regression

In summary, logistic regression allows us to classify outcomes using both continuous and categorical predictors. It provides probabilities for classification and assesses the significance of variables.

Key takeaway

Logistic regression is a powerful tool in both traditional statistics and machine learning.

Exercises

- **Exercise 1:** Explain the difference between linear regression and logistic regression in terms of the type of outcomes they predict.

Exercises

- **Exercise 1:** Explain the difference between linear regression and logistic regression in terms of the type of outcomes they predict.
- **Exercise 2:** Given a dataset with mouse weights and obesity status, calculate the probability of a mouse being obese using logistic regression.

Exercises

- **Exercise 1:** Explain the difference between linear regression and logistic regression in terms of the type of outcomes they predict.
- **Exercise 2:** Given a dataset with mouse weights and obesity status, calculate the probability of a mouse being obese using logistic regression.
- **Exercise 3:** Identify which variables might be significant predictors of obesity in a hypothetical study and justify your choices.

Exercises

- **Exercise 1:** Explain the difference between linear regression and logistic regression in terms of the type of outcomes they predict.
- **Exercise 2:** Given a dataset with mouse weights and obesity status, calculate the probability of a mouse being obese using logistic regression.
- **Exercise 3:** Identify which variables might be significant predictors of obesity in a hypothetical study and justify your choices.
- **Exercise 4:** Discuss the implications of using maximum likelihood estimation in logistic regression compared to least squares in linear regression.